DOCUMENT RESUME

ED 133 331 TM 005 561

AUTHOR Doherty, Victor W.; Forster, Fred
TITLE Can Rasch Scaled Scores Be Predicted from a Calibrated Item Pool?
PUB DATE 9 Apr 76
NOTE 7p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Correlation; *Item Banks; Prediction; Raw Scores; Scores; Statistical Analysis; Test Construction
IDENTIFIERS *Rasch Scaled Scores

ABSTRACT

In order to verify that a test made of items selected from a pool produces the same Rasch scaled achievement scores as would be predicted from previous Rasch difficulty calibrations for those items, the following experiment was conducted. Seven math tests with similar content but graduated difficulty were administered to students in grade seven. The tests overlapped each other so that each test included about 20 items that were also included in one of the other six tests. The items in these seven tests were calibrated for difficulty using the Rasch procedure. Through the linking data provided by the common items, each of the items in all of the tests was adjusted to the same scale of difficulty. These items then formed a single pool of items. These items were listed in ascending difficulty order and were divided into seven level tests each with 30 items. These new tests were administered to about 1500 students in the seventh grade. Rasch item calibrations were again computed, and linking data used to adjust items to a common achievement scale. Rasch scaled achievement scores were computed for each possible raw score for each of the level tests and for the same items as calibrated in the original pool. The comparable scaled achievement scores for the tests were established on the basis of the actual performance of students on the second administration. Tables present the results. (RC)

Can Rasch Scaled Scores Be Predicted From A Calibrated Item Pool?

Victor W. Doherty and Fred Forster

Portland Public Schools

One of the more promising uses of the Rasch model is the support it can provide in constructing a statistically sound item pool. If the items in the pool are screened for consistency in level of difficulty and discrimination, the information provided by the Rasch procedure makes it possible to choose any subset of items from that pool to form a test and develop known statistical characteristics based on previous experience with the items. In addition to providing a range of scores which can be used to determine the rank order performance of students, the Rasch procedure can provide a measure for each student on a performance continuum which spans the full range of the item bank. In addition to providing an interval level estimate of ability for each raw score, the Rasch procedure provides a standard error of the estimate for each score. By developing a single underlying scale, it is relatively straightforward to relate the performance of different groups to other groups (normative comparisons).

We have verified experimentally that the Rasch procedure produces an equal-interval scale of very high quality and objectivity. We have also verified that as long as items represent a consistent content area and are R calibrated, we can achieve reliable estimates of achievement level for all sets of items drawn from the pool.

In order to verify that a test made of items selected from a pool produces the same Rasch scaled achievement scores as would be predicted

from previous Rasch difficulty calibrations for those items, we conducted the following experiment.

In the spring of 1974, seven math tests with similar content but graduated difficulty were administered to approximately 1500 students in grade 7. The tests overlapped each other so that each test included about 20 items that were also included in one of the other six tests. These seven tests were designated as forms W, X, Y, Z, D1, D2, and D3. The items in these seven tests were calibrated for difficulty using the Rasch procedure. Through the linking data provided by the common items, each of the items in all of the tests was adjusted to the same scale of difficulty. These items then formed a single pool of items.

Items in this pool were listed in ascending difficulty order. They were then divided into seven level tests each with 30 items (about 13 overlapping up, 12 down, and 6 unique to each test). In the spring of 1975, these new tests were administered to about 1500 students in the seventh grade. Rasch calibrations were again computed, and linking data used to adjust items to a common achievement scale.

Rasch scaled achievement scores were computed for each possible raw score for each of the level tests administered in spring, 1975, and for the same items as calibrated in the original pool administered in spring, 1974. It is important to remember that the original calibrations in 1974 were based on different groups taking different tests from those used in spring, 1975. The comparable scaled achievement scores for the 1975 tests were established on the basis of the actual performance of students on the second (1975) administration.

The results are shown in Tables 1-7. Note that for Level Test #1, the number of students taking the easiest level was too small to produce reliable results (N=22). Tables 1-7 compare Rasch scaled achievement scores for Level Tests 2-7, derived from calibrations based on the administration of these level tests, with the scaled scores derived from calibrations for the same items as administered in the tests that formed the original item pool. Also shown are the mean (arbitrarily set at 50) and standard deviation of the score distributions, significance of difference between means, correlations, and discrepancy between the level test scaled achievement scores and the corresponding score for the same pool items.

Figures 1-7 show the correlation plots, again indicating the very close correspondence of scaled achievement scores from the two administrations.

Insert tables, figures

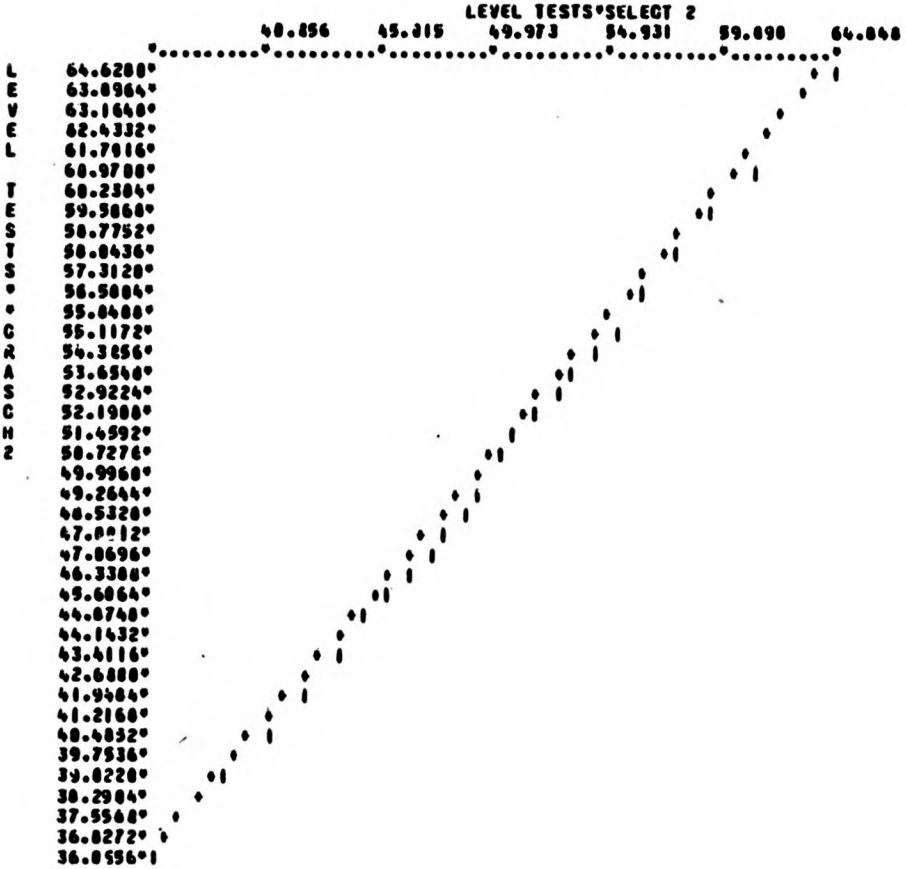
N = 280

LINKING LEVEL TESTS PREDICTED TO LEVEL TESTS ACTUAL 08/11/75

MEAN TEST 1= 50.00 MEAN TEST 2= 50.00

STAN. DEVIATION 1= 7.2165 STAN. DEVIATION 2= 7.3642 DIFFERENCE BETWEEN THE MEANS= .0811 CORRELATION= .9789

SCORE	ACHIEVEMENT1	ACHIEVEMENT2	ACH. EST.	DISCREPANCY	Z-DIFFERENCE
1	35.364	35.098	35.365	.267	.041
2	38.718	38.481	38.719	.238	.050
3	40.777	40.568	40.778	.210	.053
4	42.312	42.130	42.313	.183	.052
5	43.566	43.409	43.567	.158	.049
6	44.645	44.514	44.646	.132	.043
7	45.609	45.502	45.610	.108	.037
8	46.494	46.411	46.495	.084	.030
9	47.324	47.263	47.325	.062	.023
10	48.114	48.076	48.115	.039	.015
11	48.879	48.863	48.880	.017	.006
12	49.630	49.634	49.631	-.003	-.001
13	50.376	50.401	50.377	-.024	-.009
14	51.126	51.171	51.127	-.044	-.017
15	51.891	51.955	51.892	-.063	-.024
16	52.681	52.764	52.682	-.082	-.030
17	53.509	53.610	53.510	-.100	-.036
18	54.393	54.512	54.394	-.118	-.041
19	55.356	55.493	55.357	-.136	-.045
20	56.424	56.588	56.435	-.153	-.047
21	57.625	57.856	57.686	-.170	-.048
22	59.219	59.406	59.220	-.186	-.047
23	61.276	61.480	61.277	-.203	-.043
24	64.628	64.848	6[illegible]29	-.219	-.033


LEVEL TESTS*SELECT 2
40.856
45.915
49.973
54.931
59.890
64.848
L
E
V
E
L
T
E
S
T
S
*
*
C
R
A
S
C
H
2
64.6288*
63.8964*
63.1640*
62.4332*
61.7016*
60.9708*
60.2384*
59.5068*
58.7752*
58.0436*
57.3120*
56.5804*
55.8488*
55.1172*
54.3856*
53.6540*
52.9224*
52.1908*
51.4592*
50.7276*
49.9960*
49.2644*
48.5328*
47.8012*
47.0696*
46.3380*
45.6064*
44.8748*
44.1432*
43.4116*
42.6800*
41.9484*
41.2168*
40.4852*
39.7536*
39.0220*
38.2904*
37.5588*
36.8272*
36.0956*